Project Proposal: Apple iPhone Brand Sentiment Analysis

Project Overview and Motivation

This project aims to analyze consumer sentiment regarding Apple iPhones by leveraging Natural Language Processing (NLP) techniques. Given Apple's global brand presence, understanding consumer perceptions is vital for informed brand management and targeted improvements. Using the approaches outlined in Sruthy Sugunan's work, this project will explore sentiment classification using machine learning models enhanced by genetic algorithm optimization to maximize accuracy and applicability.

Dataset and Data Source

We will utilize review datasets for Apple iPhones from sources like Amazon and Trustpilot, collected via web scraping. This diverse dataset will allow for robust sentiment analysis on various models of the iPhone. We plan to preprocess these datasets for consistency and build a unified sentiment dataset categorized by polarity.

NLP Methods and Models

The project will implement:

Rule-Based Models: Naive Bayes and Logistic Regression for baseline sentiment classification.

Advanced Classifiers: Support Vector Machine (SVM) as the primary model due to its effectiveness in text-based classification, optimized further with TPOT (Tree-based Pipeline Optimization Tool) to explore hyperparameters and enhance performance.

Interpretability and Explainability: LIME (Local Interpretable Model-agnostic Explanations) to provide insight into model predictions.

Packages and Tools

Data Collection and Processing: BeautifulSoup and requests for web scraping; pandas and NLTK for data cleaning, tokenization, and stopword removal.

Model Training: scikit-learn for machine learning models; TPOT for genetic algorithm-based optimization.

Performance Evaluation: Metrics will include accuracy, F1 score, precision, and recall, with detailed results visualized in ROC curves.

Tasks and Schedule

Data Collection and Preprocessing (Weeks 1-2): Web scraping, data cleaning, and text normalization.

Model Implementation (Weeks 3-4): Build and evaluate Naive Bayes, Logistic Regression, and SVM models.

Optimization and Testing (Week 5): Use TPOT to optimize hyperparameters, assess model improvements, and interpret results with LIME.

Final Report and Presentation Preparation (Weeks 7-8): Compile findings, draft report, and prepare presentation slides.

Evaluation Metrics

The model performance will be measured using accuracy, precision, recall, and F1 score, with a specific focus on the SVM model optimized through genetic algorithms. Comparative analysis with baseline models will also be performed to determine improvements due to optimization.

References

Sugunan, S. (2023). Enhancing Apple Product Review Sentiment Analysis using Machine Learning and Genetic Algorithm Optimization. Dublin Business School.